



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification : C12N 15/31, A61K 39/04, C07K 14/35, C12N 15/62, A61K 38/16, G01N 33/569, C12Q 1/68, C07K 16/12</p>	A1	<p>(11) International Publication Number: WO 98/44119</p> <p>(43) International Publication Date: 8 October 1998 (08.10.98)</p>											
<p>(21) International Application Number: PCT/DK98/00132</p> <p>(22) International Filing Date: 1 April 1998 (01.04.98)</p> <p>(30) Priority Data:</p> <table border="0"> <tr> <td>0376/97</td> <td>2 April 1997 (02.04.97)</td> <td>DK</td> </tr> <tr> <td>60/044,624</td> <td>18 April 1997 (18.04.97)</td> <td>US</td> </tr> <tr> <td>1277/97</td> <td>10 November 1997 (10.11.97)</td> <td>DK</td> </tr> <tr> <td>69/070,488</td> <td>5 January 1998 (05.01.98)</td> <td>US</td> </tr> </table>	0376/97	2 April 1997 (02.04.97)	DK	60/044,624	18 April 1997 (18.04.97)	US	1277/97	10 November 1997 (10.11.97)	DK	69/070,488	5 January 1998 (05.01.98)	US	<p>Walter [FF/DK]; Ingemannsvej 21, DK-1964 Frederiksberg C (DK).</p> <p>(74) Agent: PLOUGMANN, VINGTOFT & PARTNERS A/S; Sankt Anne Plads 11, P.O. Box 3007, DK-1021 Copenhagen K (DK).</p> <p>(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UZ, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, HE, IE, LI, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published</p> <p>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>
0376/97	2 April 1997 (02.04.97)	DK											
60/044,624	18 April 1997 (18.04.97)	US											
1277/97	10 November 1997 (10.11.97)	DK											
69/070,488	5 January 1998 (05.01.98)	US											
<p>(71) Applicant (for all designated States except US): STATENS SERUM INSTITUT [DK/DK], Artillerivej 5, DK-2300 Copenhagen S (DK).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): ANDERSEN, Peter [DK/DK]; Lystrupvej 7, DK-2700 Brønshøj (DK); NIELSEN, Rikke [DK/DK]; Tjørvaldsensvej 9, 5. sal, DK-1871 Frederiksberg C (DK); ROSENKRANDS, Ida [DK/DK]; Ragnhildsgade 70, 1. th, DK-2100 Copenhagen Ø (DK); WELDINGH, Karin [DK/DK]; Nørrebrogade 224, 3. tv, DK-2400 Copenhagen N (DK); RASMUSSEN, Peter, Birk [DK/DK]; Rudolph Berghsgade 5, DK-2100 Copenhagen Ø (DK); ØHTTINGER, Thomas [DK/DK]; Egebjerg Allé 12, DK-2900 Hellerup (DK); FLORIO,</p>													
<p>(54) Title: NUCLEIC ACID FRAGMENTS AND POLYPEPTIDE FRAGMENTS DERIVED FROM <i>M. TUBERCULOSIS</i></p>													
<p>(57) Abstract</p> <p>The present invention is based on the identification and characterization of a number of <i>M. tuberculosis</i> derived novel proteins and protein fragments (SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, and 169-171). The invention is directed to the polypeptides and immunologically active fragments thereof, the genes encoding them, immunological compositions such as vaccines and skin test reagents containing the polypeptides. Another part of the invention is based on the surprising discovery that fusions between ESAT-6 and MPT59 are superior immunogens compared to each of the unfused proteins, respectively.</p>													

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LI	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BD	Bhutan	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Ivory Coast	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BT	Bhutan	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Cote d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroun		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CZ	Czech Republic	KZ	Kazakhstan	RO	Romania		
DE	Germany	LC	Saint Lucia	RU	Russian Federation		
DK	Denmark	LI	Liechtenstein	SD	Sudan		
EE	Estonia	LB	Lebanon	SE	Sweden		
		LR	Liberia	SG	Singapore		

NUCLEIC ACID FRAGMENTS AND POLYPEPTIDE FRAGMENTS DERIVED FROM
M. TUBERCULOSIS

FIELD OF THE INVENTION

The present invention relates to a number of immunologically
5 active, novel polypeptide fragments derived from the *Mycobac-*
terium tuberculosis, vaccines and other immunologic composi-
tions containing the fragments as immunogenic components, and
methods of production and use of the polypeptides. The inven-
tion also relates to novel nucleic acid fragments derived
10 from *M. tuberculosis* which are useful in the preparation of
the polypeptide fragments of the invention or in the diag-
nosis of infection with *M. tuberculosis*. The invention fur-
ther relates to certain fusion polypeptides, notably fusions
between ESAT-6 and MPT59.

15 BACKGROUND OF THE INVENTION

Human tuberculosis (hereinafter designated "TB") caused by
Mycobacterium tuberculosis is a severe global health problem
responsible for approximately 3 million deaths annually,
according to the WHO. The worldwide incidence of new TB cases
20 has been progressively falling for the last decade but the
recent years has markedly changed this trend due to the
advent of AIDS and the appearance of multidrug resistant
strains of *M. tuberculosis*.

The only vaccine presently available for clinical use is BCG.
25 a vaccine which efficacy remains a matter of controversy. BCG
generally induces a high level of acquired resistance in
animal models of TB, but several human trials in developing
countries have failed to demonstrate significant protection.
Notably, BCG is not approved by the FDA for use in the United
30 States.

This makes the development of a new and improved vaccine
against TB an urgent matter which has been given a very high

- priority by the WHO. Many attempts to define protective mycobacterial substances have been made, and from 1950 to 1970 several investigators reported an increased resistance after experimental vaccination. However, the demonstration of
- 5 a specific long-term protective immune response with the potency of BCG has not yet been achieved by administration of soluble proteins or cell wall fragments, although progress is currently being made by relying on polypeptides derived from short term-culture filtrate, cf. the discussion below.
- 10 Immunity to *M. tuberculosis* is characterized by three basic features; i) Living bacilli efficiently induces a protective immune response in contrast to killed preparations; ii) Specifically sensitized T lymphocytes mediate this protection; iii) The most important mediator molecule seems to be
- 15 interferon gamma (INF- γ).

- Short term-culture filtrate (ST-CP) is a complex mixture of proteins released from *M. tuberculosis* during the first few days of growth in a liquid medium (Andersen et al., 1991). Culture filtrates has been suggested to hold protective
- 20 antigens recognized by the host in the first phase of TB infection (Andersen et al. 1991, Orme et al. 1993). Recent data from several laboratories have demonstrated that experimental subunit vaccines based on culture filtrate antigens can provide high levels of acquired resistance to TB (Pal and
- 25 Horwitz, 1992; Roberts et al., 1995; Andersen, 1994; Lindblad et al., 1997). Culture filtrates are, however, complex protein mixtures and until now very limited information has been available on the molecules responsible for this protective immune response. In this regard, only two culture filtrate
- 30 antigens have been described as involved in protective immunity, the low mass antigen ESAT-6 (Andersen et al., 1995 and EP-A-0 706 571) and the 31 kDa molecule Ag85E (EP-0 432 203).

There is therefore a need for the identification of further antigens involved in the induction of protective immunity

against TB in order to eventually produce an effective sub-unit vaccine.

OBJECT OF THE INVENTION

It is an object of the invention to provide novel antigens which are effective as components in a subunit vaccine against TB or which are useful as components in diagnostic compositions for the detection of infection with mycobacteria, especially virulence-associated mycobacteria. The novel antigens may also be important drug targets.

10 SUMMARY OF THE INVENTION

The present invention is i.a. based on the identification and characterization of a number of previously uncharacterized culture filtrate antigens from *M. tuberculosis*. In animal models of TB, T cells mediating immunity are focused predominantly to antigens in the regions 6-12 and 17-30 kDa of SF-CF. In the present invention 9 antigens in the low molecular weight region (CFP7, CFP7A, CFP7B, CFP8A, CFP8B, CFP9, CFP10A, and CFP11) and 18 antigens (CFP16, CFP17, CFP19, CFP19B, CFP20, CFP21, CFP22, CFP22A, CFP23, CFP23A, CFP23B, CFP25, CFP26, CFP27, CFP28, CFP29, CFP30A, and CFP30B) in the 17-30 kDa region have been identified. Of these, CFP19A and CFP23 have been selected because they exhibit relatively high homologies with CFP21 and CFP25, respectively, in so far that a nucleotide homology sequence search in the Sanger Database (cf. below) with the genes encoding CFP21 and CFP25, (cfp25 and cfp21 respectively), shows homology to two *M. tuberculosis* DNA sequences, orf19A and orf23. The two sequences, orf19A and orf23, encode to putative proteins CFP19A and CFP23 with the molecular weights of approx. 19 and 23 kDa respectively. The identity, at amino acid level, to CFP21 and CFP25 is 46% and 50%, respectively, for both proteins. CFP21 and CFP25 have been shown to be dominant T-cell antigens, and it is therefore believed that CFP19A and CFP23 are possible new T-cell antigens.

Furthermore, a 50 kDa antigen (CFP50) has been isolated from culture filtrate and so has also an antigen (CWP32) isolated from the cell wall in the 30 kDa region.

5 The present invention is also based on the identification of a number of putative antigens from *M. tuberculosis* which are not present in *Mycobacterium bovis* BCG strains. The nucleotide sequences encoding these putative antigens are: rdi-orf2, rdi-orf3, rdi-orf4, rdi-orf5, rdi-orf8, rdi-orf9a, and rdi-orf9b.

10 Finally, the invention is based on the surprising discovery that fusions between ESAT-6 and MPT59 are superior immunogens compared to the unfused proteins, respectively.

15 The encoding genes for 33 of the antigens have been determined, the distribution of a number of the antigens in various mycobacterial strains investigated and the biological activity of the products characterized. The panel hold antigens with potential for vaccine purposes as well as for diagnostic purposes, since the antigens are all secreted by metabolizing mycobacteria.

20 The following table lists the antigens of the invention by the names used herein as well as by reference to relevant SEQ ID NOs of N-terminal sequences, full amino acid sequences and sequences of DNA encoding the antigens:

	Antigen	N-terminal sequence SEQ ID NO:	Nucleotide sequence SEQ ID NO:	Amino acid sequence SEQ ID NO:
	CFP7		1	2
	CFP7A	81	47	48
	CFP7B	168	146	147
5	CFP8A	73	148	149
	CFP8B	74	150	151
	CFP9		3	4
	CFP10A	169	140	141
	CFP11	170	142	143
10	CFP16	79	53	54
	CFP17	17	5	6
	CFP19	82	49	50
	CFP19A		51	52
	CFP19B	80		
15	CFP20	18	7	8
	CFP21	19	9	10
	CFP22	20	11	12
	CFP22A	83	53	54
	CFP23		55	56
20	CFP23A	76		
	CFP23B	75		
	CFP25	21	13	14
	CFP25A	78	65	66
	CFP27	84	57	58
25	CFP28	22		
	CFP29	23	15	16
	CFP30A	85	59	60
	CFP30B	171	144	145
	CFP50	86	61	62
30	MPT51		41	42
	CWF32	77	152	153
	RD1-ORF8		67	68
	RD1-ORF2		71	72
	RD1-ORF9B		69	70
35	RD1-ORF3		87	88
	RD1-ORF9A		93	94
	RD1-ORF4		89	90
	RD1-ORF5		91	92
40	MPT59- ESAT6			172
	ESAT6- MPT59			173

It is well-known in the art that T-cell epitopes are respon-
sible for the elicitation of the acquired immunity against
45 TB, whereas B-cell epitopes are without any significant
influence on acquired immunity and recognition of mycobacte-
ria *in vivo*. Since such T-cell epitopes are linear and are
known to have a minimum length of 6 amino acid residues, the

present invention is especially concerned with the identification and utilisation of such T-cell epitopes.

Hence, in its broadest aspect the invention relates to a substantially pure polypeptide fragment which

- 5 a) comprises an amino acid sequence selected from the sequences shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, and any one of 168-
10 171,
- b) comprises a subsequence of the polypeptide fragment defined in a) which has a length of at least 6 amino acid residues, said subsequence being immunologically equivalent to the polypeptide defined in a) with
15 respect to the ability of evoking a protective immune response against infections with mycobacteria belonging to the tuberculosis complex or with respect to the ability of eliciting a diagnostically significant immune response indicating previous or ongoing sensitization with antigens derived from mycobacteria belonging
20 to the tuberculosis complex, or
- c) comprises an amino acid sequence having a sequence identity with the polypeptide defined in a) or the subsequence defined in b) of at least 70% and at the
25 same time being immunologically equivalent to the polypeptide defined in a) with respect to the ability of evoking a protective immune response against infections with mycobacteria belonging to the tuberculosis complex or with respect to the ability of eliciting a diagnostically significant immune response indicating
30 previous or ongoing sensitization with antigens derived from mycobacteria belonging to the tuberculosis complex,

with the proviso that

i) the polypeptide fragment is in essentially pure form when consisting of the amino acid sequence 1-96 of SEQ ID NO: 2 or when consisting of the amino acid sequence 87-108 of SEQ ID NO: 4 fused to β -galactosidase,

ii) the degree of sequence identity in c) is at least 95% when the polypeptide comprises a homologue of a polypeptide which has the amino acid sequence SEQ ID NO: 12 or a subsequence thereof as defined in b), and

iii) the polypeptide fragment contains a threonine residue corresponding to position 213 in SEQ ID NO: 42 when comprising an amino acid sequence of at least 6 amino acids in SEQ ID NO: 42.

Other parts of the invention pertains to the DNA fragments encoding a polypeptide with the above definition as well as to DNA fragments useful for determining the presence of DNA encoding such polypeptides.

DETAILED DISCLOSURE OF THE INVENTION

In the present specification and claims, the term "polypeptide fragment" denotes both short peptides with a length of at least two amino acid residues and at most 10 amino acid residues, oligopeptides (11-100 amino acid residues), and longer peptides (the usual interpretation of "polypeptide", i.e. more than 100 amino acid residues in length) as well as proteins (the functional entity comprising at least one peptide, oligopeptide, or polypeptide which may be chemically modified by being glycosylated, by being lipidated, or by comprising prosthetic groups). The definition of polypeptides also comprises native forms of peptides/proteins in mycobacteria as well as recombinant proteins or peptides in any type of expression vectors transforming any kind of host, and also chemically synthesized peptides.

In the present context the term "substantially pure polypeptide fragment" means a polypeptide preparation which contains at most 5% by weight of other polypeptide material with which it is natively associated (lower percentages of other polypeptide material are preferred, e.g. at most 4%, at most 3%, at most 2%, at most 1%, and at most $\frac{1}{2}\%$). It is preferred that the substantially pure polypeptide is at least 96% pure, i.e. that the polypeptide constitutes at least 96% by weight of total polypeptide material present in the preparation, and higher percentages are preferred, such as at least 97%, at least 98%, at least 99%, at least 99.25%, at least 99.5%, and at least 99.75%. It is especially preferred that the polypeptide fragment is in "essentially pure form", i.e. that the polypeptide fragment is essentially free of any other antigen with which it is natively associated, i.e. free of any other antigen from bacteria belonging to the tuberculosis complex. This can be accomplished by preparing the polypeptide fragment by means of recombinant methods in a non-mycobacterial host cell as will be described in detail below, or by synthesizing the polypeptide fragment by the well-known methods of solid or liquid phase peptide synthesis, e.g. by the method described by Merrifield or variations thereof.

The term "subsequence" when used in connection with a polypeptide of the invention having a SEQ ID NO selected from 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, and any one of 168-171 denotes any continuous stretch of at least 6 amino acid residues taken from the *M. tuberculosis* derived polypeptides in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 and being immunological equivalent thereto with respect to the ability of conferring increased resistance to infections with bacteria belonging to the tuberculosis complex. Thus, included is also a

polypeptide from different sources, such as other bacteria or even from eukaryotic cells.

When referring to an "immunologically equivalent" polypeptide is herein meant that the polypeptide, when formulated in a vaccine or a diagnostic agent (i.e. together with a pharmaceutically acceptable carrier or vehicle and optionally an adjuvant), will

- I) confer, upon administration (either alone or as an immunologically active constituent together with other antigens), an acquired increased specific resistance in a mouse and/or in a guinea pig and/or in a primate such as a human being against infections with bacteria belonging to the tuberculosis complex which is at least 20% of the acquired increased resistance conferred by *Mycobacterium bovis* BCG and also at least 20% of the acquired increased resistance conferred by the parent polypeptide comprising SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 (said parent polypeptide having substantially the same relative location and pattern in a 2DE gel prepared as the 2DE gel shown in Fig. 6, cf. the examples), the acquired increased resistance being assessed by the observed reduction in mycobacterial counts from spleen, lung or other organ homogenates isolated from the mouse or guinea pig receiving a challenge infection with a virulent strain of *M. tuberculosis*, or, in a primate such as a human being, being assessed by determining the protection against development of clinical tuberculosis in a vaccinated group versus that observed in a control group receiving a placebo or BCG (preferably the increased resistance is higher and corresponds to at least 50% of the protective immune response elicited by *M. bovis* BCG, such as at least 60%, or even more preferred to at least 80% of

the protective immune response elicited by *M. bovis* BCG, such as at least 90%; in some cases it is expected that the increased resistance will supersede that conferred by *M. bovis* BCG, and hence it is preferred that the resistance will be at least 100%, such as at least 110% of said increased resistance); and/or

- II) elicit a diagnostically significant immune response in a mammal indicating previous or ongoing sensitization with antigens derived from mycobacteria belonging to the tuberculosis complex; this diagnostically significant immune response can be in the form of a delayed type hypersensitivity reaction which can e.g. be determined by a skin test, or can be in the form of IFN- γ release determined e.g. by an IFN- γ assay as described in detail below. A diagnostically significant response in a skin test setup will be a reaction which gives rise to a skin reaction which is at least 5 mm in diameter and which is at least 65% (preferably at least 75% such as at the least 85%) of the skin reaction (assessed as the skin reaction diameter) elicited by the parent polypeptide comprising SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171.

The ability of the polypeptide fragment to confer increased immunity may thus be assessed by measuring in an experimental animal, e.g. a mouse or a guinea pig, the reduction in mycobacterial counts from the spleen, lung or other organ homogenates isolated from the experimental animal which have received a challenge infection with a virulent strain of mycobacteria belonging to the tuberculosis complex after previously having been immunized with the polypeptide, as compared to the mycobacterial counts in a control group of experimental animals infected with the same virulent strain, which experimental animals have not previously been immunized

against tuberculosis. The comparison of the mycobacterial counts may also be carried out with mycobacterial counts from a group of experimental animals receiving a challenge infection with the same virulent strain after having been immunized with *Mycobacterium bovis* BCG.

The mycobacterial counts in homogenates from the experimental animals immunized with a polypeptide fragment according to the present invention must at the most be 5 times the counts in the mice or guinea pigs immunized with *Mycobacterium bovis* BCG, such as at the most 3 times the counts, and preferably at the most 2 times the counts.

A more relevant assessment of the ability of the polypeptide fragment of the invention to confer increased resistance is to compare the incidence of clinical tuberculosis in two groups of individuals (e.g. humans or other primates) where one group receives a vaccine as described herein which contains an antigen of the invention and the other group receives either a placebo or an other known TB vaccine (e.g. BCG). In such a setup, the antigen of the invention should give rise to a protective immunity which is significantly higher than the one provided by the administration of the placebo (as determined by statistical methods known to the skilled artisan).

The "tuberculosis-complex" has its usual meaning, i.e. the complex of mycobacteria causing TB which are *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, and *Mycobacterium africanum*.

In the present context the term "metabolizing mycobacteria" means live mycobacteria that are multiplying logarithmically and releasing polypeptides into the culture medium wherein they are cultured.

The term "sequence identity" indicates a quantitative measure of the degree of homology between two amino acid sequences or

between two nucleotide sequences of equal length: The sequence identity can be calculated as $\frac{(N_{ref}-N_{dif})100}{N_{ref}}$, wherein

N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of
5 residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC ($N_{dif}=2$ and $N_{ref}=8$).

The sequence identity is used here to illustrate the degree of identity between the amino acid sequence of a given
10 polypeptide and the amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171. The amino acid sequence to be compared with the amino
15 acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 may be deduced from a DNA sequence, e.g. obtained by hybridization as defined
20 below, or may be obtained by conventional amino acid sequencing methods. The sequence identity is preferably determined on the amino acid sequence of a mature polypeptide, i.e. without taking any leader sequence into consideration.

As appears from the above disclosure, polypeptides which are
25 not identical to the polypeptides having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 are embraced by the present invention. The invention allows for
30 minor variations which do not have an adverse effect on immunogenicity compared to the parent sequences and which may give interesting and useful novel binding properties or biological functions and immunogenicities etc.

Each polypeptide fragment may thus be characterized by specific amino acid and nucleic acid sequences. It will be understood that such sequences include analogues and variants produced by recombinant methods wherein such nucleic acid and polypeptide sequences have been modified by substitution, insertion, addition and/or deletion of one or more nucleotides in said nucleic acid sequences to cause the substitution, insertion, addition or deletion of one or more amino acid residues in the recombinant polypeptide. When the term DNA is used in the following, it should be understood that for the number of purposes where DNA can be substituted with RNA, the term DNA should be read to include RNA embodiments which will be apparent for the man skilled in the art. For the purposes of hybridization, PNA may be used instead of DNA, as PNA has been shown to exhibit a very dynamic hybridization profile (PNA is described in Nielsen P E et al., 1991, Science 254: 1497-1500).

In both immunodiagnosics and vaccine preparation, it is often possible and practical to prepare antigens from segments of a known immunogenic protein or polypeptide. Certain epitopic regions may be used to produce responses similar to those produced by the entire antigenic polypeptide. Potential antigenic or immunogenic regions may be identified by any of a number of approaches, e.g., Jameson-Wolf or Kyte-Doolittle antigenicity analyses or Hopp and Woods (1981) hydrophobicity analysis (see, e.g., Jameson and Wolf, 1988; Kyte and Doolittle, 1982; or U.S. Patent No. 4,554,101). Hydrophobicity analysis assigns average hydrophilicity values to each amino acid residue from these values average hydrophilicities can be calculated and regions of greatest hydrophilicity determined. Using one or more of these methods, regions of predicted antigenicity may be derived from the amino acid sequence assigned to the polypeptides of the invention.

Alternatively, in order to identify relevant T-cell epitopes which are recognized during an immune response, it is also possible to use a "brute force" method: Since T-cell epitopes

are linear, deletion mutants of polypeptides having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 will, if constructed systematically, reveal what regions of the polypeptides are essential in immune recognition, e.g. by subjecting these deletion mutants to the IFN- γ assay described herein. Another method utilises overlapping oligomers (preferably synthetic having a length of e.g. 20 amino acid residues) derived from polypeptides having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171. Some of these will give a positive response in the IFN- γ assay whereas others will not.

In a preferred embodiment of the invention, the polypeptide fragment of the invention comprises an epitope for a T-helper cell.

Although the minimum length of a T-cell epitope has been shown to be at least 6 amino acids, it is normal that such epitopes are constituted of longer stretches of amino acids. Hence it is preferred that the polypeptide fragment of the invention has a length of at least 7 amino acid residues, such as at least 8, at least 9, at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 22, at least 24, and at least 30 amino acid residues.

As will appear from the examples, a number of the polypeptides of the invention are natively translation products which include a leader sequence (or other short peptide sequences), whereas the product which can be isolated from short-term culture filtrates from bacteria belonging to the tuberculosis complex are free of these sequences. Although it may in some applications be advantageous to produce these polypeptides recombinantly and in this connection facilitate export of the polypeptides from the host cell by including

information encoding the leader sequence in the gene for the polypeptide, it is more often preferred to either substitute the leader sequence with one which has been shown to be superior in the host system for effecting export, or to
5 totally omit the leader sequence (e.g. when producing the polypeptide by peptide synthesis. Hence, a preferred embodiment of the invention is a polypeptide which is free from amino acid residues -39 to -1 in SEQ ID NO: 6 and/or -32 to -1 in SEQ ID NO: 10 and/or -8 to -1 in SEQ ID NO: 12 and/or
10 -32 to -1 in SEQ ID NO: 14 and/or -33 to -1 in SEQ ID NO: 42 and/or -38 to -1 in SEQ ID NO: 52 and/or -33 to -1 in SEQ ID NO: 56 and/or -56 to -1 in SEQ ID NO: 58 and/or -28 to -1 in SEQ ID NO: 151.

In another preferred embodiment, the polypeptide fragment of
15 the invention is free from any signal sequence; this is especially interesting when the polypeptide fragment is produced synthetically but even when the polypeptide fragments are produced recombinantly it is normally acceptable that they are not exported by the host cell to the periplasm
20 or the extracellular space; the polypeptide fragments can be recovered by traditional methods (cf. the discussion below) from the cytoplasm after disruption of the host cells, and if there is need for refolding of the polypeptide fragments, general refolding schemes can be employed, cf. e.g. the
25 disclosure in WO 94/18227 where such a general applicable refolding method is described.

A suitable assay for the potential utility of a given polypeptide fragment derived from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60,
30 62, 64, 66, 68, 70, any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171 is to assess the ability of the polypeptide fragment to effect IFN- γ release from primed memory T-lymphocytes. Polypeptide fragments which have this capability are according to the
35 invention especially interesting embodiments of the invention: It is contemplated that polypeptide fragments which

stimulate T lymphocyte immune response shortly after the onset of the infection are important in the control of the mycobacteria causing the infection before the mycobacteria have succeeded in multiplying up to the number of bacteria that would have resulted in fulminant infection.

Thus, an important embodiment of the invention is a polypeptide fragment defined above which

- 1) induces a release of IFN- γ from primed memory T-lymphocytes withdrawn from a mouse within 2 weeks of primary infection or within 4 days after the mouse has been re-challenge infected with mycobacteria belonging to the tuberculosis complex, the induction performed by the addition of the polypeptide to a suspension comprising about 200,000 spleen cells per ml, the addition of the polypeptide resulting in a concentration of 1-4 μ g polypeptide per ml suspension, the release of IFN- γ being assessable by determination of IFN- γ in supernatant harvested 2 days after the addition of the polypeptide to the suspension, and/or
- 2) induces a release of IFN- γ of at least 1,500 pg/ml above background level from about 1,000,000 human PBMC (peripheral blood mononuclear cells) per ml isolated from TB patients in the first phase of infection, or from healthy BCG vaccinated donors, or from healthy contacts to TB patients, the induction being performed by the addition of the polypeptide to a suspension comprising the about 1,000,000 PBMC per ml, the addition of the polypeptide resulting in a concentration of 1-4 μ g polypeptide per ml suspension, the release of IFN- γ being assessable by determination of IFN- γ in supernatant harvested 2 days after the addition of the polypeptide to the suspension; and/or
- 3) induces an IFN- γ release from bovine PBMC derived from animals previously sensitized with mycobacteria belong-

ing to the tuberculosis complex, said release being at least two times the release observed from bovine PBMC derived from animals not previously sensitized with mycobacteria belonging to the tuberculosis complex.

- 5 Preferably, in alternatives 1 and 2, the release effected by the polypeptide fragment gives rise to at least 1,500 pg/ml IFN- γ in the supernatant but higher concentrations are preferred, e.g. at least 2,000 pg/ml and even at least 3,000 pg/ml IFN- γ in the supernatant. The IFN- γ release from bovine
10 PBMC can e.g. be measured as the optical density (OD) index over background in a standard cytokine ELISA and should thus be at least two, but higher numbers such as at least 3, 5, 8, and 10 are preferred.

- The polypeptide fragments of the invention preferably comprises an amino acid sequence of at least 6 amino acid residues in length which has a higher sequence identity than 70 percent with SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, any one of 17-23, 42, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70,
15 any one of 72-86, 88, 90, 92, 94, 141, 143, 145, 147, 149, 151, 153, or any one of 168-171. A preferred minimum percentage of sequence identity is at least 80%, such as at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, and at least 99.5%.

- 25 As mentioned above, it will normally be interesting to omit the leader sequences from the polypeptide fragments of the invention. However, by producing fusion polypeptides, superior characteristics of the polypeptide fragments of the invention can be achieved. For instance, fusion partners
30 which facilitate export of the polypeptide when produced recombinantly, fusion partners which facilitate purification of the polypeptide, and fusion partners which enhance the immunogenicity of the polypeptide fragment of the invention are all interesting possibilities. Therefore, the invention
35 also pertains to a fusion polypeptide comprising at least one

- polypeptide fragment defined above and at least one fusion partner. The fusion partner can, in order to enhance immunogenicity, e.g. be selected from the group consisting of another polypeptide fragment as defined above (so as to allow for multiple expression of relevant epitopes), and an other polypeptide derived from a bacterium belonging to the tuberculosis complex, such as ESAT-6, MPB64, MPT64, and MPB59 or at least one T-cell epitope of any of these antigens. Other immunogenicity enhancing polypeptides which could serve as fusion partners are T-cell epitopes (e.g. derived from the polypeptides ESAT-6, MPB64, MPT64, or MPB59) or other immunogenic epitopes enhancing the immunogenicity of the target gene product, e.g. lymphokines such as INF- γ , IL-2 and IL-12. In order to facilitate expression and/or purification the fusion partner can e.g. be a bacterial fimbrial protein, e.g. the pilus components pilin and papA; protein A; the ZZ-peptide (ZZ-fusions are marketed by Pharmacia in Sweden); the maltose binding protein; glutathione S-transferase; β -galactosidase; or poly-histidine.
- Other interesting fusion partners are polypeptides which are lipidated and thereby effect that the immunogenic polypeptide is presented in a suitable manner to the immune system. This effect is e.g. known from vaccines based on the *Borrelia burgdorferi* OspA polypeptide, wherein the lipidated membrane anchor in the polypeptide confers a self-adjuvating effect to the polypeptide (which is natively lipidated) when isolated from cells producing it. In contrast, the OspA polypeptide is relatively silent immunologically when prepared without the lipidation anchor.
- As evidenced in Example 6A, the fusion polypeptide consisting of MPT59 fused directly N-terminally to ESAT-6 enhances the immunogenicity of ESAT-6 beyond what would be expected from the immunogenicities of MPT59 and ESAT-6 alone. The precise reason for this surprising finding is not yet known, but it is expected that either the presence of both antigens lead to a synergistic effect with respect to immunogenicity or the

presence of a sequence N-terminally to the ESAT-6 sequence protects this immune dominant protein from loss of important epitopes known to be present in the N-terminus. A third, alternative, possibility is that the presence of a sequence
5 C-terminally to the MPT59 sequence enhances the immunologic properties of this antigen.

Hence, one part of the invention pertains to a fusion polypeptide fragment which comprises a first amino acid sequence including at least one stretch of amino acids constituting a T-cell epitope derived from the *M. tuberculosis*
10 protein ESAT-6 or MPT59, and a second amino acid sequence including at least one T-cell epitope derived from a *M. tuberculosis* protein different from ESAT-6 (if the first stretch of amino acids are derived from ESAT-6) or MPT59 (if
15 the first stretch of amino acids are derived from MPT59) and/or including a stretch of amino acids which protects the first amino acid sequence from *in vivo* degradation or post-translational processing. The first amino acid sequence may be situated N- or C-terminally to the second amino acid
20 sequence, but in line with the above considerations regarding protection of the ESAT-6 N-terminus it is preferred that the first amino acid sequence is C-terminal to the second when the first amino acid sequence is derived from ESAT-6.

Although only the effect of fusion between MPT59 and ESAT6
25 has been investigated at present, it is believed that ESAT6 and MPT59 or epitopes derived therefrom could be advantageously be fused to other fusion partners having substantially the same effect on overall immunogenicity of the fusion construct. Hence, it is preferred that such a fusion
30 polypeptide fragment according of the invention is one, wherein the at least one T-cell epitope included in the second amino acid sequence is derived from a *M. tuberculosis* polypeptide (the "parent" polypeptide) selected from the group consisting of a polypeptide fragment according to the
35 present invention and described in detail above and in the examples, or the amino acid sequence could be derived from

any one of the *M. tuberculosis* proteins DnaK, GroEL, urease, glutamine synthetase, the proline rich complex, L-alanine dehydrogenase, phosphate binding protein, Ag 85 complex, HBHA (heparin binding hemagglutinin), MPT51, MPT64, superoxide
5 diamutase, 19 kDa lipoprotein, α -crystallin, GroES, MPT59 (when the first amino acid sequence is derived from ESAT-6), and ESAT-6 (when the first amino acid sequence is derived from MPT59). It is preferred that the first and second T-cell epitopes each have a sequence identity of at least 70% with
10 the natively occurring sequence in the proteins from which they are derived and it is even further preferred that the first and/or second amino acid sequence has a sequence identity of at least 70% with the protein from which they are derived. A most preferred embodiment of this fusion
15 polypeptide is one wherein the first amino acid sequence is the amino acid sequence of ESAT-6 or MPT59 and/or the second amino acid sequence is the full-length amino acid sequence of the possible "parent" polypeptides listed above.

In the most preferred embodiment, the fusion polypeptide
20 fragment comprises ESAT-6 fused to MPT59 (advantageously, ESAT-6 is fused to the C-terminus of MPT59) and in one special embodiment, there are no linkers introduced between the two amino acid sequences constituting the two parent polypeptide fragments.

25 Another part of the invention pertains to a nucleic acid fragment in isolated form which

- 1) comprises a nucleic acid sequence which encodes a polypeptide or fusion polypeptide as defined above, or
30 comprises a nucleic acid sequence complementary thereto, and/or
- 2) has a length of at least 10 nucleotides and hybridizes readily under stringent hybridization conditions (as defined in the art, i.e. 5-10°C under the melting point T_m , cf. Sambrook et al. 1989, pages 11.45-11.49) with a

nucleic acid fragment which has a nucleotide sequence selected from

- SEQ ID NO: 1 or a sequence complementary thereto,
SEQ ID NO: 3 or a sequence complementary thereto,
5 SEQ ID NO: 5 or a sequence complementary thereto,
SEQ ID NO: 7 or a sequence complementary thereto,
SEQ ID NO: 9 or a sequence complementary thereto,
SEQ ID NO: 11 or a sequence complementary thereto,
SEQ ID NO: 13 or a sequence complementary thereto,
10 SEQ ID NO: 15 or a sequence complementary thereto,
SEQ ID NO: 41 or a sequence complementary thereto,
SEQ ID NO: 47 or a sequence complementary thereto,
SEQ ID NO: 49 or a sequence complementary thereto,
SEQ ID NO: 51 or a sequence complementary thereto,
15 SEQ ID NO: 53 or a sequence complementary thereto,
SEQ ID NO: 55 or a sequence complementary thereto,
SEQ ID NO: 57 or a sequence complementary thereto,
SEQ ID NO: 59 or a sequence complementary thereto,
SEQ ID NO: 61 or a sequence complementary thereto,
20 SEQ ID NO: 63 or a sequence complementary thereto,
SEQ ID NO: 65 or a sequence complementary thereto,
SEQ ID NO: 67 or a sequence complementary thereto,
SEQ ID NO: 69 or a sequence complementary thereto,
SEQ ID NO: 71 or a sequence complementary thereto,
25 SEQ ID NO: 87 or a sequence complementary thereto,
SEQ ID NO: 89 or a sequence complementary thereto,
SEQ ID NO: 91 or a sequence complementary thereto,
SEQ ID NO: 93 or a sequence complementary thereto,
SEQ ID NO: 140 or a sequence complementary thereto,
30 SEQ ID NO: 142 or a sequence complementary thereto,
SEQ ID NO: 144 or a sequence complementary thereto,
SEQ ID NO: 146 or a sequence complementary thereto,
SEQ ID NO: 148 or a sequence complementary thereto,
SEQ ID NO: 150 or a sequence complementary thereto, and
35 SEQ ID NO: 152 or a sequence complementary thereto,

with the proviso that when the nucleic acid fragment comprises a subsequence of SEQ ID NO: 41, then the nucleic acid

fragment contains an A corresponding to position 781 in SEQ ID NO: 41 and when the nucleic acid fragment comprises a subsequence of a nucleotide sequence exactly complementary to SEQ ID NO: 41, then the nucleic acid fragment comprises a T
5 corresponding to position 781 in SEQ ID NO: 41.

It is preferred that the nucleic acid fragment is a DNA fragment.

To provide certainty of the advantages in accordance with the invention, the preferred nucleic acid sequence when employed
10 for hybridization studies or assays includes sequences that are complementary to at least a 10 to 40, or so, nucleotide stretch of the selected sequence. A size of at least 10 nucleotides in length helps to ensure that the fragment will be of sufficient length to form a duplex molecule that is
15 both stable and selective. Molecules having complementary sequences over stretches greater than 10 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained.

20 Hence, the term "subsequence" when used in connection with the nucleic acid fragments of the invention is intended to indicate a continuous stretch of at least 10 nucleotides exhibits the above hybridization pattern. Normally this will require a minimum sequence identity of at least 70% with a
25 subsequence of the hybridization partner having SEQ ID NO: 1, 3, 5, 7, 9, 11, 12, 15, 21, 41, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 87, 89, 91, 93, 140, 142, 144, 146, 148, 150, or 152. It is preferred that the nucleic acid fragment is longer than 10 nucleotides, such as at least 15,
30 at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, and at least 80 nucleotides long, and the sequence identity should preferably also be higher than 70%, such as at least 75%, at least 80%, at least 85%, at
35 least 90%, at least 92%, at least 94%, at least 96%, and at

least 98%. It is most preferred that the sequence identity is 100%. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR technology of U.S. Patent 4,603,102, or by introducing selected sequences into recombinant vectors for recombinant production.

It is well known that the same amino acid may be encoded by various codons, the codon usage being related, *inter alia*, to the preference of the organisms in question expressing the nucleotide sequence. Thus, at least one nucleotide or codon of a nucleic acid fragment of the invention may be exchanged by others which, when expressed, result in a polypeptide identical or substantially identical to the polypeptide encoded by the nucleic acid fragment in question. The invention thus allows for variations in the sequence such as substitution, insertion (including introns), addition, deletion and rearrangement of one or more nucleotides, which variations do not have any substantial effect on the polypeptide encoded by the nucleic acid fragment or a subsequence thereof. The term "substitution" is intended to mean the replacement of one or more nucleotides in the full nucleotide sequence with one or more different nucleotides, "addition" is understood to mean the addition of one or more nucleotides at either end of the full nucleotide sequence, "insertion" is intended to mean the introduction of one or more nucleotides within the full nucleotide sequence, "deletion" is intended to indicate that one or more nucleotides have been deleted from the full nucleotide sequence whether at either end of the sequence or at any suitable point within it, and "rearrangement" is intended to mean that two or more nucleotide residues have been exchanged with each other.

The nucleotide sequence to be modified may be of cDNA or genomic origin as discussed above, but may also be of synthetic origin. Furthermore, the sequence may be of mixed cDNA and genomic, mixed cDNA and synthetic or genomic and syn-

thetic origin as discussed above. The sequence may have been modified, e.g. by site-directed mutagenesis, to result in the desired nucleic acid fragment encoding the desired polypeptide. The following discussion focused on modifications of nucleic acid encoding the polypeptide should be understood to encompass also such possibilities, as well as the possibility of building up the nucleic acid by ligation of two or more DNA fragments to obtain the desired nucleic acid fragment, and combinations of the above-mentioned principles.

The nucleotide sequence may be modified using any suitable technique which results in the production of a nucleic acid fragment encoding a polypeptide of the invention.

The modification of the nucleotide sequence encoding the amino acid sequence of the polypeptide of the invention should be one which does not impair the immunological function of the resulting polypeptide.

A preferred method of preparing variants of the antigens disclosed herein is site-directed mutagenesis. This technique is useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, derived from the antigen sequences, through specific mutagenesis of the underlying nucleic acid. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the nucleic acid. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the nucleotide sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucle-

otides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art as exemplified by publications (Adelman et al., 1983). As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage (Messing et al., 1981). These phage are readily commercially available and their use is generally well known to those skilled in the art.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector which includes within its sequence a nucleic acid sequence which encodes the polypeptides of the invention. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically, for example by the method of Crea et al. (1978). This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as E. coli polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as E. coli cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected nucleic acid fragments of the invention using site-directed mutagenesis is provided as a means of producing potentially useful species of the genes and is not meant to be limiting as there are other ways in which sequence variants of the nucleic acid fragments of the invention may be obtained. For example, recombinant vectors encoding the desired genes may be treated with mutagenic agents to obtain sequence variants (see, e.g.,

a method described by Hichenlaub, 1979) for the mutagenesis of plasmid DNA using hydroxylamine.

The invention also relates to a replicable expression vector which comprises a nucleic acid fragment defined above, especially a vector which comprises a nucleic acid fragment
5 encoding a polypeptide fragment of the invention.

The vector may be any vector which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to
10 be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication; examples of such a vector are a plasmid, phage, cosmid, mini-chromosome or virus. Alternatively,
15 the vector may be one which, when introduced in a host cell, is integrated in the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

Expression vectors may be constructed to include any of the
20 DNA segments disclosed herein. Such DNA might encode an antigenic protein specific for virulent strains of mycobacteria or even hybridization probes for detecting mycobacteria nucleic acids in samples. Longer or shorter DNA segments could be used, depending on the antigenic protein desired.
25 Epitopic regions of the proteins expressed or encoded by the disclosed DNA could be included as relatively short segments of DNA. A wide variety of expression vectors is possible including, for example, DNA segments encoding reporter gene products useful for identification of heterologous gene
30 products and/or resistance genes such as antibiotic resistance genes which may be useful in identifying transformed cells.

The vector of the invention may be used to transform cells so as to allow propagation of the nucleic acid fragments of the

invention or so as to allow expression of the polypeptide fragments of the invention. Hence, the invention also pertains to a transformed cell harbouring at least one such vector according to the invention, said cell being one which does not natively harbour the vector and/or the nucleic acid fragment of the invention contained therein. Such a transformed cell (which is also a part of the invention) may be any suitable bacterial host cell or any other type of cell such as a unicellular eukaryotic organism, a fungus or yeast, or a cell derived from a multicellular organism, e.g. an animal or a plant. It is especially in cases where glycosylation is desired that a mammalian cell is used, although glycosylation of proteins is a rare event in prokaryotes. Normally, however, a prokaryotic cell is preferred such as a bacterium belonging to the genera *Mycobacterium*, *Salmonella*, *Pseudomonas*, *Bacillus* and *Escherichia*. It is preferred that the transformed cell is an *E. coli*, *B. subtilis*, or *M. bovis* BCG cell, and it is especially preferred that the transformed cell expresses a polypeptide according to the invention. The latter opens for the possibility to produce the polypeptide of the invention by simply recovering it from the culture containing the transformed cell. In the most preferred embodiment of this part of the invention the transformed cell is *Mycobacterium bovis* BCG strain: Danish 1331, which is the *Mycobacterium bovis* strain Copenhagen from the Copenhagen BCG Laboratory, Statens Seruminstitut, Denmark.

The nucleic acid fragments of the invention allow for the recombinant production of the polypeptides fragments of the invention. However, also isolation from the natural source is a way of providing the polypeptide fragments as is peptide synthesis.

Therefore, the invention also pertains to a method for the preparation of a polypeptide fragment of the invention, said method comprising inserting a nucleic acid fragment as defined above into a vector which is able to replicate in a

host cell, introducing the resulting recombinant vector into the host cell (transformed cells may be selected using various techniques, including screening by differential hybridization, identification of fused reporter gene products, resistance markers, anti-antigen antibodies and the like).
5 culturing the host cell in a culture medium under conditions sufficient to effect expression of the polypeptide (of course the cell may be cultivated under conditions appropriate to the circumstances, and if DNA is desired, replication conditions are used), and recovering the polypeptide from the host
10 cell or culture medium; or

isolating the polypeptide from a short-term culture filtrate as defined in claim 1; or

isolating the polypeptide from whole mycobacteria of the
15 tuberculosis complex or from lysates or fractions thereof, e.g. cell wall containing fractions, or

synthesizing the polypeptide by solid or liquid phase peptide synthesis.

The medium used to grow the transformed cells may be any
20 conventional medium suitable for the purpose. A suitable vector may be any of the vectors described above, and an appropriate host cell may be any of the cell types listed above. The methods employed to construct the vector and effect introduction thereof into the host cell may be any
25 methods known for such purposes within the field of recombinant DNA. In the following a more detailed description of the possibilities will be given:

In general, of course, prokaryotes are preferred for the initial cloning of nucleic sequences of the invention and
30 constructing the vectors useful in the invention. For example, in addition to the particular strains mentioned in the more specific disclosure below, one may mention by way of example, strains such as *E. coli* K12 strain 294 (ATCC No.

31446), *E. coli* B, and *E. coli* X 1776 (ATCC No. 31537). These examples are, of course, intended to be illustrative rather than limiting.

Prokaryotes are also preferred for expression. The
5 aforementioned strains, as well as *E. coli* W3110 (F-, lambda-
da-, prototrophic, ATCC No. 273325), bacilli such as *Bacillus*
subtilis, or other enterobacteriaceae such as *Salmonella*
typhimurium or *Serratia marcescens*, and various *Pseudomonas*
species may be used. Especially interesting are rapid-growing
10 mycobacteria, e.g. *M. smegmatis*, as these bacteria have a
high degree of resemblance with mycobacteria of the tubercu-
losis complex and therefore stand a good chance of reducing
the need of performing post-translational modifications of
the expression product.

15 In general, plasmid vectors containing replicon and control
sequences which are derived from species compatible with the
host cell are used in connection with these hosts. The vector
ordinarily carries a replication site, as well as marking
sequences which are capable of providing phenotypic selection
20 in transformed cells. For example, *E. coli* is typically
transformed using pBR322, a plasmid derived from an *E. coli*
species (see, e.g., Bolivar et al., 1977, Gene 2: 95). The
pBR322 plasmid contains genes for ampicillin and tetracycline
resistance and thus provides easy means for identifying
25 transformed cells. The pBR plasmid, or other microbial
plasmid or phage must also contain, or be modified to con-
tain, promoters which can be used by the microorganism for
expression.

These promoters most commonly used in recombinant DNA con-
30 struction include the B-lactamase (penicillinase) and lactose
promoter systems (Chang et al., 1978; Itakura et al., 1977;
Goeddel et al., 1979) and a tryptophan (*trp*) promoter system
(Goeddel et al., 1979; EPO Appl. Publ. No. 0036776). While
these are the most commonly used, other microbial promoters
35 have been discovered and utilized, and details concerning

their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally with plasmid vectors (Siebwenlist et al., 1980). Certain genes from prokaryotes may be expressed efficiently in *E. coli* from their own promoter sequences, precluding the need for addition of another promoter by artificial means.

After the recombinant preparation of the polypeptide according to the invention, the isolation of the polypeptide may for instance be carried out by affinity chromatography (or other conventional biochemical procedures based on chromatography), using a monoclonal antibody which substantially specifically binds the polypeptide according to the invention. Another possibility is to employ the simultaneous electroelution technique described by Andersen et al. in *J. Immunol. Methods* **161**: 29-39.

According to the invention the post-translational modifications involves lipidation, glycosylation, cleavage, or elongation of the polypeptide.

In certain aspects, the DNA sequence information provided by this invention allows for the preparation of relatively short DNA (or RNA or PNA) sequences having the ability to specifically hybridize to mycobacterial gene sequences. In these aspects, nucleic acid probes of an appropriate length are prepared based on a consideration of the relevant sequence. The ability of such nucleic acid probes to specifically hybridize to the mycobacterial gene sequences lend them particular utility in a variety of embodiments. Most importantly, the probes can be used in a variety of diagnostic assays for detecting the presence of pathogenic organisms in a given sample. However, other uses are envisioned, including the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructs.

Apart from their use as starting points for the synthesis of polypeptides of the invention and for hybridization probes (useful for direct hybridization assays or as primers in e.g. PCR or other molecular amplification methods) the nucleic acid fragments of the invention may be used for effecting *in vivo* expression of antigens, i.e. the nucleic acid fragments may be used in so-called DNA vaccines. Recent research have revealed that a DNA fragment cloned in a vector which is non-replicative in eukaryotic cells may be introduced into an animal (including a human being) by e.g. intramuscular injection or percutaneous administration (the so-called "gene gun" approach). The DNA is taken up by e.g. muscle cells and the gene of interest is expressed by a promoter which is functioning in eukaryotes, e.g. a viral promoter, and the gene product thereafter stimulates the immune system. These newly discovered methods are reviewed in Ulmer et al., 1993, which hereby is included by reference.

Hence, the invention also relates to a vaccine comprising a nucleic acid fragment according to the invention, the vaccine effecting *in vivo* expression of antigen by an animal, including a human being, to whom the vaccine has been administered, the amount of expressed antigen being effective to confer substantially increased resistance to infections with mycobacteria of the tuberculosis complex in an animal, including a human being.

The efficacy of such a "DNA vaccine" can possibly be enhanced by administering the gene encoding the expression product together with a DNA fragment encoding a polypeptide which has the capability of modulating an immune response. For instance, a gene encoding lymphokine precursors or lymphokines (e.g. IFN- γ , IL-2, or IL-12) could be administered together with the gene encoding the immunogenic protein, either by administering two separate DNA fragments or by administering both DNA fragments included in the same vector. It also is a possibility to administer DNA fragments comprising a multitude of nucleotide sequences which each encode

relevant epitopes of the polypeptides disclosed herein so as to effect a continuous sensitization of the immune system with a broad spectrum of these epitopes.

- As explained above, the polypeptide fragments of the invention are excellent candidates for vaccine constituents or for constituents in an immune diagnostic agent due to their extracellular presence in culture media containing metabolizing virulent mycobacteria belonging to the tuberculosis complex, or because of their high homologies with such extracellular antigens, or because of their absence in *M. bovis* BCG.

- Thus, another part of the invention pertains to an immunologic composition comprising a polypeptide or fusion polypeptide according to the invention. In order to ensure optimum performance of such an immunologic composition it is preferred that it comprises an immunologically and pharmaceutically acceptable carrier, vehicle or adjuvant.

- Suitable carriers are selected from the group consisting of a polymer to which the polypeptide(s) is/are bound by hydrophobic non-covalent interaction, such as a plastic, e.g. polystyrene, or a polymer to which the polypeptide(s) is/are covalently bound, such as a polysaccharide, or a polypeptide, e.g. bovine serum albumin, ovalbumin or keyhole limpet haemocyanin. Suitable vehicles are selected from the group consisting of a diluent and a suspending agent. The adjuvant is preferably selected from the group consisting of dimethyldioctadecylammonium bromide (DDA), Quil A, poly I:C, Freund's incomplete adjuvant, IFN- γ , IL-2, IL-12, monophosphoryl lipid A (MPL), and muramyl dipeptide (MDP).

- A preferred immunologic composition according to the present invention comprising at least two different polypeptide fragments, each different polypeptide fragment being a polypeptide or a fusion polypeptide defined above. It is

preferred that the immunologic composition comprises between 3-20 different polypeptide fragments or fusion polypeptides.

Such an immunologic composition may preferably be in the form of a vaccine or in the form of a skin test reagent.

- 5 In line with the above, the invention therefore also pertains to a method for producing an immunologic composition according to the invention, the method comprising preparing, synthesizing or isolating a polypeptide according to the invention, and solubilizing or dispersing the polypeptide in
10 a medium for a vaccine, and optionally adding other *M. tuberculosis* antigens and/or a carrier, vehicle and/or adjuvant substance.

- Preparation of vaccines which contain peptide sequences as active ingredients is generally well understood in the art,
15 as exemplified by U.S. Patents 4,606,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, all incorporated herein by reference. Typically, such vaccines are prepared as injectables either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension
20 in, liquid prior to injection may also be prepared. The preparation may also be emulsified. The active immunogenic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline,
25 dextrose, glycerol, ethanol, or the like, and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, or adjuvants which enhance the effectiveness of the vaccines.

- 30 The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional

binders and carriers may include, for example, polyalkalene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1-2%. Oral formulations include
5 such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release
10 formulations or powders and contain 10-95% of active ingredient, preferably 25-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include acid
15 addition salts (formed with the free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium,
20 potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, including, e.g.,
25 the capacity of the individual's immune system to mount an immune response, and the degree of protection desired. Suitable dosage ranges are of the order of several hundred
30 micrograms active ingredient per vaccination with a preferred range from about 0.1 μg to 1000 μg , such as in the range from about 1 μg to 300 μg , and especially in the range from about 10 μg to 50 μg . Suitable regimens for initial administration and booster shots are also variable but are typified by an
35 initial administration followed by subsequent inoculations or other administrations.

- The manner of application may be varied widely. Any of the conventional methods for administration of a vaccine are applicable. These are believed to include oral application on a solid physiologically acceptable base or in a physiologically acceptable dispersion, parenterally, by injection or the like. The dosage of the vaccine will depend on the route of administration and will vary according to the age of the person to be vaccinated and, to a lesser degree, the size of the person to be vaccinated.
- 5 Some of the polypeptides of the vaccine are sufficiently immunogenic in a vaccine, but for some of the others the immune response will be enhanced if the vaccine further comprises an adjuvant substance.
- 10 Various methods of achieving adjuvant effect for the vaccine include use of agents such as aluminum hydroxide or phosphate (alum), commonly used as 0.05 to 0.1 percent solution in phosphate buffered saline, admixture with synthetic polymers of sugars (Carbopol) used as 0.25 percent solution, aggregation of the protein in the vaccine by heat treatment with
- 15 temperatures ranging between 70° to 101°C for 30 second to 2 minute periods respectively. Aggregation by reactivating with pepsin treated (Fab) antibodies to albumin, mixture with bacterial cells such as *C. parvum* or endotoxins or lipopolysaccharide components of gram-negative bacteria, emulsion in
- 20 physiologically acceptable oil vehicles such as mannide mono-oleate (Aracel A) or emulsion with 20 percent solution of a perfluorocarbon (Fluorol-DA) used as a block substitute may also be employed. According to the invention DDA (dimethyldioctadecylammonium bromide) is an interesting candidate for an
- 25 adjuvant, but also Freund's complete and incomplete adjuvants as well as QuilA and RIBI are interesting possibilities. Further possibilities are monophosphoryl lipid A (MPL), and muramyl dipeptide (MDP).
- 30 Another highly interesting (and thus, preferred) possibility of achieving adjuvant effect is to employ the technique
- 35

described in Gosselin et al., 1992 (which is hereby incorporated by reference herein). In brief, the presentation of a relevant antigen such as an antigen of the present invention can be enhanced by conjugating the antigen to antibodies (or antigen binding antibody fragments) against the Fc γ receptors on monocytes/macrophages. Especially conjugates between antigen and anti-Fc γ RI have been demonstrated to enhance immunogenicity for the purposes of vaccination.

Other possibilities involve the use of immune modulating substances such as lymphokines (e.g. IFN- γ , IL-2 and IL-12) or synthetic IFN- γ inducers such as poly I:C in combination with the above-mentioned adjuvants. As discussed in example 3, it is contemplated that such mixtures of antigen and adjuvant will lead to superior vaccine formulations.

In many instances, it will be necessary to have multiple administrations of the vaccine, usually not exceeding six vaccinations, more usually not exceeding four vaccinations and preferably one or more, usually at least about three vaccinations. The vaccinations will normally be at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain the desired levels of protective immunity. The course of the immunization may be followed by in vitro proliferation assays of PBL (peripheral blood lymphocytes) co-cultured with ESAT-6 or ST-CP, and especially by measuring the levels of IFN- γ released from the primed lymphocytes. The assays may be performed using conventional labels, such as radionuclides, enzymes, fluorescers, and the like. These techniques are well known and may be found in a wide variety of patents, such as U.S. Patent Nos. 3,751,932; 4,174,364 and 3,949,064, as illustrative of these types of assays.

Due to genetic variation, different individuals may react with immune responses of varying strength to the same polypeptide. Therefore, the vaccine according to the inven-

tion may comprise several different polypeptides in order to increase the immune response. The vaccine may comprise two or more polypeptides, where all of the polypeptides are as defined above, or some but not all of the peptides may be derived from a bacterium belonging to the *M. tuberculosis* complex. In the latter example the polypeptides not necessarily fulfilling the criteria set forth above for polypeptides may either act due to their own immunogenicity or merely act as adjuvants. Examples of such interesting polypeptides are MPB64, MPT64, and MPB59, but any other substance which can be isolated from mycobacteria are possible candidates.

The vaccine may comprise 3-20 different polypeptides, such as 3-10 different polypeptides.

One reason for admixing the polypeptides of the invention with an adjuvant is to effectively activate a cellular immune response. However, this effect can also be achieved in other ways, for instance by expressing the effective antigen in a vaccine in a non-pathogenic microorganism. A well-known example of such a microorganism is *Mycobacterium bovis* BCG.

Therefore, another important aspect of the present invention is an improvement of the living BCG vaccine presently available, which is a vaccine for immunizing an animal, including a human being, against TB caused by mycobacteria belonging to the tuberculosis-complex, comprising as the effective component a microorganism, wherein one or more copies of a DNA sequence encoding a polypeptide as defined above has been incorporated into the genome of the microorganism in a manner allowing the microorganism to express and secrete the polypeptide.

In the present context the term "genome" refers to the chromosome of the microorganisms as well as extrachromosomally DNA or RNA, such as plasmids. It is, however, preferred that the DNA sequence of the present invention has been introduced into the chromosome of the non-pathogenic microorganism,

since this will prevent loss of the genetic material introduced.

It is preferred that the non-pathogenic microorganism is a bacterium, e.g. selected from the group consisting of the
5 genera *Mycobacterium*, *Salmonella*, *Pseudomonas* and *Eschericia*. It is especially preferred that the non-pathogenic microorganism is *Mycobacterium bovis* BCG, such as *Mycobacterium bovis* BCG strain: Danish 1331.

- The incorporation of one or more copies of a nucleotide
10 sequence encoding the polypeptide according to the invention in a mycobacterium from a *M. bovis* BCG strain will enhance the immunogenic effect of the BCG strain. The incorporation of more than one copy of a nucleotide sequence of the invention is contemplated to enhance the immune response even
15 more, and consequently an aspect of the invention is a vaccine wherein at least 2 copies of a DNA sequence encoding a polypeptide is incorporated in the genome of the microorganism, such as at least 5 copies. The copies of DNA sequences may either be identical encoding identical polypeptides or be
20 variants of the same DNA sequence encoding identical or homologues of a polypeptide, or in another embodiment be different DNA sequences encoding different polypeptides where at least one of the polypeptides is according to the present invention.
- 25 The living vaccine of the invention can be prepared by cultivating a transformed non-pathogenic cell according to the invention, and transferring these cells to a medium for a vaccine, and optionally adding a carrier, vehicle and/or adjuvant substance.
- 30 The invention also relates to a method of diagnosing TB caused by *Mycobacterium tuberculosis*, *Mycobacterium africanum* or *Mycobacterium bovis* in an animal, including a human being, comprising intradermally injecting, in the animal, a polypeptide according to the invention or a skin test reagent

- described above, a positive skin response at the location of injection being indicative of the animal having TB, and a negative skin response at the location of injection being indicative of the animal not having TB. A positive response is a skin reaction having a diameter of at least 5 mm, but larger reactions are preferred, such as at least 1 cm, 1.5 cm, and at least 2 cm in diameter. The composition used as the skin test reagent can be prepared in the same manner as described for the vaccines above.
- 10 In line with the disclosure above pertaining to vaccine preparation and use, the invention also pertains to a method for immunising an animal, including a human being, against TB caused by mycobacteria belonging to the tuberculosis complex, comprising administering to the animal the polypeptide of the
- 15 invention, or a vaccine composition of the invention as described above, or a living vaccine described above. Preferred routes of administration are the parenteral (such as intravenous and intraarterially), intraperitoneal, intramuscular, subcutaneous, intradermal, oral, buccal, sublingual,
- 20 nasal, rectal or transdermal route.

- The protein ESAT-6 which is present in short-term culture filtrates from mycobacteria as well as the *esat-6* gene in the mycobacterial genome has been demonstrated to have a very limited distribution in other mycobacterial strains than *M. tuberculosis*, e.g. *esat-6* is absent in both BCG and the
- 25 majority of mycobacterial species isolated from the environment, such as *M. avium* and *M. terrae*. It is believed that this is also the case for at least one of the antigens of the present invention and their genes and therefore, the diagnostic embodiments of the invention are especially well-suited for performing the diagnosis of on-going or previous infection with virulent mycobacterial strains of the tuberculosis complex, and it is contemplated that it will be possible to distinguish between 1) subjects (animal or human) which have
- 30 been previously vaccinated with e.g. BCG vaccines or subjected to antigens from non-virulent mycobacteria and 2)
- 35

subjects which have or have had active infection with virulent mycobacteria.

A number of possible diagnostic assays and methods can be envisaged:

- 5 When diagnosis of previous or ongoing infection with virulent mycobacteria is the aim, a blood sample comprising mononuclear cells (i.e. T-lymphocytes) from a patient could be contacted with a sample of one or more polypeptides of the invention. This contacting can be performed *in vitro* and a
- 10 positive reaction could e.g. be proliferation of the T-cells or release of cytokines such as γ -interferon into the extracellular phase (e.g. into a culture supernatant); a suitable *in vivo* test would be a skin test as described above. It is also conceivable to contact a serum sample from a subject to
- 15 contact with a polypeptide of the invention, the demonstration of a binding between antibodies in the serum sample and the polypeptide being indicative of previous or ongoing infection.

- The invention therefore also relates to an *in vitro* method
- 20 for diagnosing ongoing or previous sensitization in an animal or a human being with bacteria belonging to the tuberculosis complex, the method comprising providing a blood sample from the animal or human being, and contacting the sample from the animal with the polypeptide of the invention, a significant
 - 25 release into the extracellular phase of at least one cytokine by mononuclear cells in the blood sample being indicative of the animal being sensitized. By the term "significant release" is herein meant that the release of the cytokine is significantly higher than the cytokine release from a blood
 - 30 sample derived from a non-tuberculous subject (e.g. a subject which does not react in a traditional skin test for TB). Normally, a significant release is at least two times the release observed from such a sample.

Alternatively, a sample of a possibly infected organ may be contacted with an antibody raised against a polypeptide of the invention. The demonstration of the reaction by means of methods well-known in the art between the sample and the
5 antibody will be indicative of ongoing infection. It is of course also a possibility to demonstrate the presence of anti-mycobacterial antibodies in serum by contacting a serum sample from a subject with at least one of the polypeptide fragments of the invention and using well-known methods for
10 visualizing the reaction between the antibody and antigen.

Also a method of determining the presence of mycobacterial nucleic acids in an animal, including a human being, or in a sample, comprising administering a nucleic acid fragment of the invention to the animal or incubating the sample with the
15 nucleic acid fragment of the invention or a nucleic acid fragment complementary thereto, and detecting the presence of hybridized nucleic acids resulting from the incubation (by using the hybridization assays which are well-known in the art), is also included in the invention. Such a method of
20 diagnosing TB might involve the use of a composition comprising at least a part of a nucleotide sequence as defined above and detecting the presence of nucleotide sequences in a sample from the animal or human being to be tested which hybridize with the nucleic acid fragment (or a complementary
25 fragment) by the use of PCR technique.

The fact that certain of the disclosed antigens are not present in *M. bovis* BCG but are present in virulent mycobacteria point them out as interesting drug targets; the
30 antigens may constitute receptor molecules or toxins which facilitate the infection by the mycobacterium, and if such functionalities are blocked the infectivity of the mycobacterium will be diminished.

To determine particularly suitable drug targets among the antigens of the invention, the gene encoding at least one of
35 the polypeptides of the invention and the necessary control

sequences can be introduced into avirulent strains of mycobacteria (e.g. BCG) so as to determine which of the polypeptides are critical for virulence. Once particular proteins are identified as critical for/contributory to virulence, anti-mycobacterial agents can be designed

5 rationally to inhibit expression of the critical genes or to attack the critical gene products. For instance, antibodies or fragments thereof (such as Fab and (Fab')₂ fragments can be prepared against such critical polypeptides by methods

10 known in the art and thereafter used as prophylactic or therapeutic agents. Alternatively, small molecules can be screened for their ability to selectively inhibit expression of the critical gene products, e.g. using recombinant expression systems which include the gene's endogenous promoter, or

15 for their ability to directly interfere with the action of the target. These small molecules are then used as therapeutics or as prophylactic agents to inhibit mycobacterial virulence.

Alternatively, anti-mycobacterial agents which render a virulent mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium. Such a

25 "newly avirulent" mycobacterium would constitute a superb alternative to the above described modified BCG for vaccine purposes since it would be immunologically very similar to a virulent mycobacterium compared to e.g. BCG.

Finally, a monoclonal or polyclonal antibody, which is specifically reacting with a polypeptide of the invention in an immune assay, or a specific binding fragment of said antibody, is also a part of the invention. The production of such polyclonal antibodies requires that a suitable animal be immunized with the polypeptide and that these antibodies are

30 subsequently isolated, suitably by immune affinity chromatography. The production of monoclonals can be effected by

35

methods well-known in the art, since the present invention provides for adequate amounts of antigen for both immunization and screening of positive hybridomas.

LEGENDS TO THE FIGURES

- 5 Fig. 1: Long term memory immune mice are very efficiently protected towards an infection with *M. tuberculosis*. Mice were given a challenge of *M. tuberculosis* and spleens were isolated at different time points. Spleen lymphocytes were stimulated in vitro with ST-CF and the release of IFN- γ investigated (panel A). The counts of CFU in the spleens of the two groups of mice are indicated in panel B. The memory immune mice control infection within the first week and produce large quantities of IFN- γ in response to antigens in ST-CF.
- 10
- 15 Fig. 2: T cells involved in protective immunity are predominantly directed to molecules from 6-12 and 17-38 kDa. Splenic T cells were isolated four days after the challenge with *M. tuberculosis* and stimulated in vitro with narrow molecular mass fractions of ST-CF. The release of IFN- γ was investigated
- 20
- Fig. 3: Nucleotide sequence (SEQ ID NO: 1) of *cfp7*. The deduced amino acid sequence (SEQ ID NO: 2) of CFP7 is given in conventional one-letter code below the nucleotide sequence. The putative ribosome-binding site is written in underlined italics as are the putative -10 and -35 regions. Nucleotides written in bold are those encoding CFP7.
- 25
- Fig. 4: Nucleotide sequence (SEQ ID NO: 3) of *cfp9*. The deduced amino acid sequence (SEQ ID NO: 4) of CFP9 is given in conventional one-letter code below the nucleotide sequence. The putative ribosome-binding site Shine Dalgarno sequence is written in underlined italics as are the putative -10 and -35 regions. Nucleotides in bold writing are those
- 30

encoding CFP9. The nucleotide sequence obtained from the lambda 226 phage is double underlined.

Fig. 5: Nucleotide sequence of *mpt51*. The deduced amino acid sequence of MPT51 is given in a one-letter code below the nucleotide sequence. The signal is indicated in italics, the putative potential ribosome-binding site is underlined. The nucleotide difference and amino acid difference compared to the nucleotide sequence of MPB51 (Ohara et al., 1995) are underlined at position 780. The nucleotides given in italics are not present in *M. tuberculosis* H37Rv.

Fig. 6: the position of the purified antigens in the 2DE system have been determined and mapped in a reference gel. The newly purified antigens are encircled and the position of well-known proteins are also indicated.

15 EXAMPLE 1

Identification of single culture filtrate antigens involved in protective immunity

A group of efficiently protected mice was generated by infecting 8-12 weeks old female C57Bl/6j mice with 5×10^4 *M. tuberculosis* i.v. After 30 days of infection the mice were subjected to 60 days of antibiotic treatment with isoniazid and were then left for 200-240 days to ensure the establishment of resting long-term memory immunity. Such memory immune mice are very efficiently protected against a secondary infection (Fig. 1). Long lasting immunity in this model is mediated by a population of highly reactive CD4 cells recruited to the site of infection and triggered to produce large amounts of IFN- γ in response to ST-CF (Fig. 1) (Andersen et al. 1995).

30 We have used this model to identify single antigens recognized by protective T cells. Memory immune mice were reinfected with 1×10^6 *M. tuberculosis* i.v. and splenic

lymphocytes were harvested at day 4-6 of reinfection, a time point where this population is highly reactive to ST-CP. The antigens recognized by these T cells were mapped by the multi-elution technique (Andersen and Heron, 1993). This technique divides complex protein mixtures separated in SDS-PAGE into narrow fractions in a physiological buffer. These fractions were used to stimulate spleen lymphocytes in vitro and the release of IFN- γ was monitored (Fig. 2). Long-term memory immune mice did not recognize these fractions before TB infection, but splenic lymphocytes obtained during the recall of protective immunity recognized a range of culture filtrate antigens and peak production of IFN- γ was found in response to proteins of apparent molecular weight 6-12 and 17-30 kDa (Fig. 2). It is therefore concluded that culture filtrate antigens within these regions are the major targets recognized by memory effector T-cells triggered to release IFN- γ during the first phase of a protective immune response.

EXAMPLE 2

Cloning of genes expressing low mass culture filtrate antigens

In example 1 it was demonstrated that antigens in the low molecular mass fraction are recognized strongly by cells isolated from memory immune mice. Monoclonal antibodies (mAbs) to these antigens were therefore generated by immunizing with the low mass fraction in RIBI adjuvant (first and second immunization) followed by two injections with the fractions in aluminium hydroxide. Fusion and cloning of the reactive cell lines were done according to standard procedures (Kohler and Milstein 1975). The procedure resulted in the provision of two mAbs: ST-3 directed to a 9 kDa culture filtrate antigen (CFP9) and EV-2 directed to a 7 kDa antigen (CFP7), when the molecular weight is estimated from migration of the antigens in an SDS-PAGE.

In order to identify the antigens binding to the Mab's, the following experiments were carried out:

The recombinant λ gt11 *M. tuberculosis* DNA library constructed by R. Young (Young, R.A. et al. 1985) and obtained through the World Health Organization IMMTUB programme (WHO.0032.wibr) was screened for phages expressing gene products which would bind the monoclonal antibodies ST-3 and PV-2.

Approximately 1×10^5 pfu of the gene library (containing approximately 25% recombinant phages) were plated on *Escherichia coli* Y1090 (DlacU169, proA⁺, Dlon, araD139, supF, trpC22::tn10 [pMC9] ATCC#37197) in soft agar and incubated for 2,5 hours at 42°C.

The plates were overlaid with sheets of nitrocellulose saturated with isopropyl- β -D-thiogalactopyranoside and incubation was continued for 2,5 hours at 37°C. The nitrocellulose was removed and incubated with samples of the monoclonal antibodies in PBS with Tween 20 added to a final concentration of 0.05%. Bound monoclonal antibodies were visualized by horseradish peroxidase-conjugated rabbit anti-mouse immunoglobulins (P260, Dako, Glostrup, DK) and a staining reaction involving 5,5',3,3'-tetramethylbenzidine and H₂O₂.

Positive plaques were recloned and the phages originating from a single plaque were used to lysogenize *E. coli* Y1089 (DlacU169, proA⁺, Dlon, araD139, strA, hfl150 [chr::tn10] [pMC9] ATCC nr. 37196). The resultant lysogenic strains were used to propagate phage particles for DNA extraction. These lysogenic *E. coli* strains have been named:

AA226 (expressing ST-3 reactive polypeptide CFP9) which has been deposited 28 June 1993 with the collection of Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) under the accession number DSM 8377 and in accordance with the provisions of the Budapest Treaty, and

AA242 (expressing PV-2 reactive polypeptide CFP7) which has been deposited 28 June 1993 with the collection of Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) under the accession number DSM 8379 and in accordance with the provisions of the Budapest Treaty.

These two lysogenic *E. coli* strains are disclosed in WO 95/01441 as are the mycobacterial polypeptide products expressed thereby. However, no information concerning the amino acid sequences of these polypeptides or their genetic origin are given, and therefore only the direct expression products of AA226 and AA242 are made available to the public.

The st-3 binding protein is expressed as a protein fused to β -galactosidase, whereas the pv-2 binding protein appears to be expressed in an unfused version.

Sequencing of the nucleotide sequence encoding the PV-2 and ST-3 binding protein

In order to obtain the nucleotide sequence of the gene encoding the pv-2 binding protein, the approximately 3 kb *M. tuberculosis* derived EcoRI - EcoRI fragment from AA242 was subcloned in the EcoRI site in the pBluescriptSK + (Stratagene) and used to transform *E. coli* XL-1Blue (Stratagene).

Similarly, to obtain the nucleotide sequence of the gene encoding the st-3 binding protein, the approximately 5 kb *M. tuberculosis* derived EcoRI - EcoRI fragment from AA226 was subcloned in the EcoRI site in the pBluescriptSK + (Stratagene) and used to transform *E. coli* XL-1Blue (Stratagene).

The complete DNA sequence of both genes were obtained by the dideoxy chain termination method adapted for supercoiled DNA by use of the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., Cleveland, OH) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems).

according to the instructions provided. The sequences DNA are shown in SEQ ID NO: 1 (CFP7) and in SEQ ID NO: 3 (CFP9) as well as in Figs. 3 and 4, respectively. Both strands of the DNA were sequenced.

5 CFP7

An open reading frame (ORF) encoding a sequence of 96 amino acid residues was identified from an ATG start codon at position 91-93 extending to a TAG stop codon at position 379-381. The deduced amino acid sequence is shown in SEQ ID NO: 2 (and in Fig. 3 where conventional one-letter amino acid codes are used).

CFP7 appear to be expressed in *E. coli* as an unfused version. The nucleotide sequence at position 78-84 is expected to be the Shine Dalgarno sequence and the sequences from position 15 47-50 and 14-19 are expected to be the -10 and -35 regions, respectively:

CFP9

The protein recognised by ST-3 was produced as a β -galactosidase fusion protein, when expressed from the AA226 lambda 20 phage. The fusion protein had an approx. size of 116 - 117kDa (Mw for β -galactosidase 116.25 kDa) which may suggest that only part of the CFP9 gene was included in the lambda clone (AA226).

Based on the 90 bp nucleotide sequence obtained on the insert 25 from lambda phage AA226, a search of homology to the nucleotide sequence of the *M. tuberculosis* genome was performed in the Sanger database (Sanger Mycobacterium tuberculosis database):

<http://www.sanger.ac.uk/pathogens/TB-blast-server.html>;

Williams, 1996). 100% identity to the cloned sequence was found on the MTCY48 cosmid. An open reading frame (ORF) encoding a sequence of 109 amino acid residues was identified from a GTG start codon at position 141 - 143 extending to a TGA stop codon at position 465 - 467. The deduced amino acid sequence is shown in Fig. 4 using conventional one letter code.

The nucleotide sequence at position 123 - 130 is expected to be the Shine Delgarno sequence and the sequences from position 73 - 78 and 4 - 9 are expected to be the -10 and -35 region respectively (Fig. 4). The ORF overlapping with the 5'-end of the sequence of AA229 is shown in Fig. 4 by double underlining.

Subcloning CFP7 and CFP9 in expression vectors

The two ORFs encoding CFP7 and CFP9 were PCR cloned into the pMST24 (Theisen et al., 1995) expression vector pRVN01 or the pQE-32 (QIAGEN) expression vector pRVN02, respectively.

The PCR amplification was carried out in a thermal reactor (Rapid cycler, Idaho Technology, Idaho) by mixing 10 ng plasmid DNA with the mastermix (0.5 μ M of each oligonucleotide primer, 0.25 μ M BSA (Stratagene), low salt buffer (20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgSO_4 and 0.1% Triton X-100) (Stratagene), 0.25 mM of each deoxynucleoside triphosphate and 0.5 U Tag Plus Long DNA polymerase (Stratagene)). Final volume was 10 μ l (all concentrations given are concentrations in the final volume). Predenaturation was carried out at 94°C for 30 s. 30 cycles of the following was performed; Denaturation at 94°C for 20 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min.

The oligonucleotide primers were synthesised automatically on a DNA synthesizer (Applied Biosystems, Foster City, Ca, ABI-391, PCR-mode), deblocked, and purified by ethanol precipitation.

The *cfp7* oligonucleotides (TABLE 1) were synthesised on the basis of the nucleotide sequence from the CFP7 sequence (Fig. 3). The oligonucleotides were engineered to include an *SmaI* restriction enzyme site at the 5' end and a *BamHI* restriction enzyme site at the 3' end for directed subcloning.

The *cfp9* oligonucleotides (TABLE 1) were synthesized partly on the basis of the nucleotide sequence from the sequence of the AA229 clone and partly from the identical sequence found in the Sanger database cosmid MTCY48 (Fig. 4). The oligonucleotides were engineered to include a *SmaI* restriction enzyme site at the 5' end and a *HindIII* restriction enzyme site at the 3' end for directed subcloning.

CFP2

By the use of PCR a *SmaI* site was engineered immediately 5' of the first codon of the ORF of 291 bp, encoding the *cfp7* gene, so that only the coding region would be expressed, and a *BamHI* site was incorporated right after the stop codon at the 3' end. The 291 bp PCR fragment was cleaved by *SmaI* and *BamHI*, purified from an agarose gel and subcloned into the *SmaI* - *BamHI* sites of the pMST24 expression vector. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue (pRVN01).

CFP9

By the use of PCR a *SmaI* site was engineered immediately 5' of the first codon of an ORF of 327 bp, encoding the *cfp9* gene, so that only the coding region would be expressed, and a *HindIII* site was incorporated after the stop codon at the 3' end. The 327 bp PCR fragment was cleaved by *SmaI* and *HindIII*, purified from an agarose gel, and subcloned into the *SmaI* - *HindIII* sites of the pQE-32 (QIAGEN) expression vector. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue (pRVN02).

Purification of recombinant CFP7 and CFP9

- The ORFs were fused N-terminally to the (His)₆-tag (cf. EP-A-G 282 242). Recombinant antigen was prepared as follows: Briefly, a single colony of *E. coli* harbouring either the
- 5 pRVN01 or the pRVN02 plasmid, was inoculated into Luria-Bertani broth containing 100 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to OD_{600nm} = 0.5. IPTG (isopropyl-β-D-thiogalactoside) was then added to a final concentration of 2 mM (expression was regulated either by the
- 10 strong IPTG inducible P_{tac} or the T5 promoter) and growth was continued for further 2 hours. The cells were harvested by centrifugation at 4,200 x g at 4°C for 8 min. The pelleted bacteria were stored overnight at -20°C. The pellet was resuspended in BC 40/100 buffer (20 mM Tris-HCl pH 7.9, 20%
- 15 glycerol, 100 mM KCl, 40 mM Imidazole) and cells were broken by sonication (5 times for 30 s with intervals of 30 s) at 4°C, followed by centrifugation at 12,000 x g for 30 min at 4°C, the supernatant (crude extract) was used for purification of the recombinant antigens.
- 20 The two Histidine fusion proteins (His-rCFP7 and His-rCFP9) were purified from the crude extract by affinity chromatography on a Ni²⁺-NTA column from QIAGEN with a volume of 100 ml. His-rCFP7 and His-rCFP9 binds to Ni²⁺. After extensive washes of the column in BC 40/100 buffer, the fusion protein
- 25 was eluted with a BC 1000/100 buffer containing 100 mM imidazole, 20 mM Tris pH 7.9, 20% glycerol and 1 M KCl. subsequently, the purified products were dialysed extensively against 10 mM Tris pH 8.0. His-rCFP7 and His-rCFP9 were then separated from contaminants by fast protein liquid chromatography (FPLC) over an anion-exchange column (Mono Q, Pharmacia, Sweden). in 10 mM Tris pH 8.0 with a linear gradient of NaCl from 0 to 1 M. Aliquots of the fractions were analyzed by 10%-20% gradient sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Fractions containing purified
- 30 either purified His-rCFP7 or His-rCFP9 were pooled.
- 35

TABLE 1. Sequence of the *cfp7* and *cfp9* oligonucleotides^a.

Orientation and oligonucleotide	Sequences (5' → 3')	Position ^b (nucleotide)
Sense		
5 pvr3	<u>GCACACCCCGGATGTCGCRAATCATG</u> (SEQ ID NO: 43)	91-105 (SEQ ID NO: 1)
str2	<u>CTAACACCCCGGGTGGCCGCCGACCG</u> (SEQ ID NO: 44)	141-155 (SEQ ID NO: 3)
Antisense		
pvr4	<u>CTACTAAGCTTGATCCCTAGCCGCCCATTTGGCGG</u> (SEQ ID NO: 45)	381-362 (SEQ ID NO: 1)
str3	<u>CTACTAAGCTTCCATGGTTCAGGTCTTTTGGATGCTTAC</u> (SEQ ID NO: 46)	467 - 447 (SEQ ID NO: 3)

- 10 ^a The *cfp7* oligonucleotides were based on the nucleotide sequence shown in Fig. 3 (SEQ ID NO: 1). The *cfp9* oligonucleotides were based on the nucleotide sequence shown in Fig. 4 (SEQ ID NO: 3). Nucleotides underlined are not contained in the nucleotide sequence of *cfp7* and *cfp9*.
- 15 ^b The positions referred to are of the non-underlined part of the primers and correspond to the nucleotide sequence shown in Fig. 3 and Fig. 4, respectively.

EXAMPLE 2A

- Identification of antigens which are not expressed in BCG strains.
- 20 In an effort to control the treat of TB, attenuated bacillus Calmette-Guérin (BCG) has been used as a live attenuated vaccine. BCG is an attenuated derivative of a virulent *Mycobacterium bovis*. The original BCG from the Pasteur Institute in Paris, France was developed from 1908 to 1921 by 231 passages in liquid culture and has never been shown to revert to virulence in animals, indicating that the attenuating mutation(s) in BCG are stable deletions and/or multiple mutations which do not readily revert. While physiological differences between BCG and *M. tuberculosis* and *M. bovis* has been noted, the attenuating mutations which arose during serial passage of the original BCG strain has been unknown until recently. The first mutations described are the loss of the gene encoding MPB64 in some BCG strains (Li et al., 1993, 25 Cettlinger and Andersen, 1994) and the gene encoding ESAT-6 in all BCG strain tested (Harboe et al., 1996). later 3 large deletions in BCG have been identified (Mahairas et al., 1996). The region named RD1 includes the gene encoding ESAT-6
- 35

and an other (RD2) the gene encoding MPT64. Both antigens have been shown to have diagnostic potential and ESAT-6 has been shown to have properties as a vaccine candidate (cf. PCT/DK94/00273 and PCT/DK/00270). In order to find new *M. tuberculosis* specific diagnostic antigens as well as antigens for a new vaccine against TB, the RD1 region (17.499 bp) of *M. tuberculosis* H37Rv has been analyzed for Open Reading Frames (ORF). ORFs with a minimum length of 96 bp have been predicted using the algorithm described by Borodovsky and McIninch (1993), in total 27 ORFs have been predicted, 20 of these have possible diagnostic and/or vaccine potential, as they are deleted from all known BCG strains. The predicted ORFs include ESAT-6 (RD1-ORF7) and CFP10 (RD1-ORF6) described previously (Sørensen et al., 1995), as a positive control for the ability of the algorithm. In the present is described the potential of 7 of the predicted antigens for diagnosis of TB as well as potential as candidates for a new vaccine against TB.

Seven open reading frames (ORF) from the 17,499kb RD1 region (Accession no. U34848) with possible diagnostic and vaccine potential have been identified and cloned.

Identification of the ORF's rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf6, rd1-orf9a, and rd1-orf9b.

The nucleotide sequence of rd1-orf2 from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 71. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 72.

The nucleotide sequence of rd1-orf3 from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 87. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 88.

The nucleotide sequence of rd1-orf4 from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 89. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 90.

The nucleotide sequence of *rdl-orf5* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 91. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 92.

- The nucleotide sequence of *rdl-orf8* from *M. tuberculosis*
5 H37Rv is set forth in SEQ ID NO: 67. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 68.

The nucleotide sequence of *rdl-orf9a* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 93. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 94.

- 10 The nucleotide sequence of *rdl-orf9b* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 69. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 76.

- The DNA sequence *rdl-orf2* (SEQ ID NO: 71) contained an open reading frame starting with an ATG codon at position 889 -
15 991 and ending with a termination codon (TAA) at position 2662 - 2664 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 72) contains 591 residues corresponding to a molecular weight of 64,525.

- 20 The DNA sequence *rdl-orf3* (SEQ ID NO: 87) contained an open reading frame starting with an ATG codon at position 2807 - 2809 and ending with a termination codon (TAA) at position 3101 - 3103 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 88) con-
25 tains 98 residues corresponding to a molecular weight of 9,799.

- The DNA sequence *rdl-orf4* (SEQ ID NO: 89) contained an open reading frame starting with a GTG codon at position 4014 - 4012 and ending with a termination codon (TAG) at position
30 3597 - 3595 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 90) con-

tains 139 residues corresponding to a molecular weight of 14,210.

The DNA sequence rd1-orf5 (SEQ ID NO: 91) contained an open reading frame starting with a GTG codon at position 3128 - 3130 and ending with a termination codon (TGA) at position 4241 - 4243 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 92) contains 371 residues corresponding to a molecular weight of 37,647.

10 The DNA sequence rd1-orf8 (SEQ ID NO: 67) contained an open reading frame starting with a GTG codon at position 5502 - 5506 and ending with a termination codon (TAG) at position 5084 - 5082 (position numbers referring to the location in RD1), and the deduced amino acid sequence (SEQ ID NO: 68)
15 contains 139 residues with a molecular weight of 11,737.

The DNA sequence rd1-orf9a (SEQ ID NO: 93) contained an open reading frame starting with a GTG codon at position 6146 - 6148 and ending with a termination codon (TAA) at position 7070 - 7072 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 94) contains 308 residues corresponding to a molecular weight of 33,453.

The DNA sequence rd1-orf9b (SEQ ID NO: 69) contained an open reading frame starting with an ATG codon at position 5072 - 5074 and ending with a termination codon (TAA) at position 7070 - 7072 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 70) contains 666 residues corresponding to a molecular weight of 70,650.

Cloning of the ORF's *rdl-orf2*, *rdl-orf3*, *rdl-orf4*, *rdl-orf5*,
rdl-orf8, *rdl-orf9a*, and *rdl-orf9b*.

The ORF's *rdl-orf2*, *rdl-orf3*, *rdl-orf4*, *rdl-orf5*, *rdl-orf8*,
rdl-orf9a and *rdl-orf9b* were PCR cloned in the pMST24 (Thei-
5 sen et al., 1995) (*rdl-orf3*) or the pQE32 (QIAGEN) (*rdl-orf2*,
rdl-orf4, *rdl-orf5*, *rdl-orf8*, *rdl-orf9a* and *rdl-orf9b*) ex-
pression vector. Preparation of oligonucleotides and PCR
amplification of the *rdl-orf* encoding genes, was carried out
as described in example 2. Chromosomal DNA from *M. tuberculo-*
10 *sis* H37Rv was used as template in the PCR reactions. Oligonu-
cleotides were synthesized on the basis of the nucleotide
sequence from the RDI region (Accession no. U34848). The
oligonucleotide primers were engineered to include an re-
striction enzyme site at the 5' end and at the 3' end by
15 which a later subcloning was possible. Primers are listed in
TABLE 2.

rdl-orf2. A *Bam*HI site was engineered immediately 5' of the
first codon of *rdl-orf2*, and a *Hind*III site was incorporated
right after the stop codon at the 3' end. The gene *rdl-orf2*
20 was subcloned in pQE32, giving pT096.

rdl-orf3. A *Sma*I site was engineered immediately 5' of the
first codon of *rdl-orf3*, and a *Nco*I site was incorporated
right after the stop codon at the 3' end. The gene *rdl-orf3*
was subcloned in pMST24, giving pT087.

25 *rdl-orf4*. A *Bam*HI site was engineered immediately 5' of the
first codon of *rdl-orf4*, and a *Hind*III site was incorporated
right after the stop codon at the 3' end. The gene *rdl-orf4*
was subcloned in pQE32, giving pT089.

30 *rdl-orf5*. A *Bam*HI site was engineered immediately 5' of the
first codon of *rdl-orf5*, and a *Hind*III site was incorporated
right after the stop codon at the 3' end. The gene *rdl-orf5*
was subcloned in pQE32, giving pT088.

rdl-orf8. A *Bam*HI site was engineered immediately 5' of the first codon of *rdl-orf8*, and a *Nco*I site was incorporated right after the stop codon at the 3' end. The gene *rdl-orf8* was subcloned in pMST24, giving pTO98.

- 5 *rdl-orf9a*. A *Bam*HI site was engineered immediately 5' of the first codon of *rdl-orf9a*, and a *Hind*III site was incorporated right after the stop codon at the 3' end. The gene *rdl-orf9a* was subcloned in pQE32, giving pTO91.

- 10 *rdl-orf9b*. A *Sca*I site was engineered immediately 5' of the first codon of *rdl-orf9b*, and a *Hind* III site was incorporated right after the stop codon at the 3' end. The gene *rdl-orf9b* was subcloned in pQE32, giving pTO96.

- The PCR fragments were digested with the suitable restriction
15 enzymes, purified from an agarose gel and cloned into either pMST24 or pQE-32. The seven constructs were used to transform the *E. coli* XL1-Blue. Endpoints of the gene fusions were determined by the dideoxy chain termination method. Both strands of the DNA were sequenced.

- 20 Purification of recombinant RD1-ORF2, RD1-ORF3, RD1-ORF4, RD1-ORF5, RD1-ORF8, RD1-ORF9a and RD1-ORF9b.

- The rRD1-ORFs were fused N-terminally to the (His)₆ -tag. Recombinant antigen was prepared as described in example 2 (with the exception that pTO91 was expressed at 30°C and not
25 at 37°C), using a single colony of *E. coli* harbouring either the pTO87, pTO88, pTO89, pTO90, pTO91, pTO96 or pTO98 for inoculation. Purification of recombinant antigen by Ni²⁺ affinity chromatography was also carried out as described in example 2. Fractions containing purified His-rRD1-ORF2, His-
30 rRD1-ORF3 His-rRD1-ORF4, His-rRD1-ORF5, His-rRD1-ORF8, His-rRD1-ORF9a or His-rRD1-ORF9b were pooled. The His-rRD1-ORF's were extensively dialysed against 10 mM Tris/HCl, pH 8.5, 3 M urea followed by an additional purification step performed on an anion exchange column (Mono Q) using fast protein liquid

chromatography (FPLC) (Pharmacia, Uppsala, Sweden). The purification was carried out in 10 mM Tris/HCl, pH 8.5, 3 M urea and protein was eluted by a linear gradient of NaCl from 0 to 1 M. Fractions containing the His-rD1-ORF's were pooled and subsequently dialysed extensively against 25 mM Hepes, pH 8.0 before use.

Table 2. Sequence of the rdl-orf's oligonucleotides*.

Orientation and oligonucleotide	Sequences (5' → 3')	Position (nt)
10 Sense		
RD1-ORF2f	<u>CTGGGATCCGCATGACTGCTGAACCG</u>	886 - 903
RD1-ORF3f	<u>CTTCCGGGATCGAATAATGTCAC</u>	2807 - 2822
RD1-ORF4f	<u>GTAGGATCCTAGGAGACATCAGCGC</u>	4028 - 4015
RD1-ORF5f	<u>CTGGGATCCCGGTGATCACCATGCTGTGG</u>	3028 - 3045
15 RD1-ORF6f	<u>CTCGGATCCGTTGGGTGCAGGTCGGCGATGGGC</u>	5502 - 5479
RD1-ORF9af	<u>GTTGATGTGAGCTCAGGTGAAGAAGGTGAAG</u>	6144 - 6160
RD1-ORF9bf	<u>GTTGATGTGAGCTCTATGGCGGCGAGTACAC</u>	5072 - 5089
Antisense		
RD1-ORF2r	<u>TGCAAGCTTTTAACCGGCGCTTGGGGTTC</u>	2664 - 2644
20 RD1-ORF3r	<u>GATGCCAGGTTAGCGGAAGACGCGGC</u>	3103 - 3086
RD1-ORF4r	<u>CGATCTAGCTTGGCATTGAGGCTTA</u>	3582 - 3597
RD1-ORF5r	<u>TGCGAGCTTTTACCAGTGTGCTCTTGTTC</u>	4243 - 4223
RD1-ORF6r	<u>CTCCCATGCTACGACAAAGTCTTCCGGCGGC</u>	5083 - 5105
RD1-ORF9a/dr	<u>CGATCTAGCTTTTCAACGACGTCACGCC</u>	7073 - 7056

- 25 * The oligonucleotides were constructed from the Accession number U34484 nucleotide sequence (Mahairas et al., 1996). Nucleotides (nt) underlined are not contained in the nucleotide sequence of RD1-ORF's. The positions correspond to the nucleotide sequence of Accession number U34484.

The nucleotide sequences of rdl-orf2, rdl-orf3, rdl-orf4, rdl-orf5, rdl-orf8, rdl-orf9a, and rdl-orf9b from *M. tuberculosis* H37Rv are set forth in SEQ ID NO: 71, 87, 89, 91, 67, 93, and 69, respectively. The deduced amino acid sequences of rdl-orf2, rdl-orf3, rdl-orf4, rdl-orf5, rdl-orf8, rdl-orf9a, and rdl-orf9b are set forth in SEQ ID NO: 72, 88, 90, 92, 68, 94, and 70, respectively.